

## Research Article

# Prediction of Soil Organic Carbon for Ethiopian Highlands Using Soil Spectroscopy

**Tadele Amare,<sup>1,2</sup> Christian Hergarten,<sup>1</sup> Hans Hurni,<sup>1</sup> Bettina Wolfgramm,<sup>1</sup> Birru Yitaferu,<sup>2</sup> and Yihenew G. Selassie<sup>3</sup>**

<sup>1</sup> Centre for Development and Environment, University of Bern, Hallerstrasse 10, CH-3012 Bern, Switzerland

<sup>2</sup> Amhara Regional Agricultural Research Institute, P.O. Box 527, Bahir Dar, Ethiopia

<sup>3</sup> Bahir Dar University, P.O. Box 79, Bahir Dar, Ethiopia

Correspondence should be addressed to Tadele Amare; [tadele17b@yahoo.com](mailto:tadele17b@yahoo.com)

Received 30 April 2013; Accepted 30 May 2013

Academic Editors: M. Bernoux and G. Broll

Copyright © 2013 Tadele Amare et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soil spectroscopy was applied for predicting soil organic carbon (SOC) in the highlands of Ethiopia. Soil samples were acquired from Ethiopia's National Soil Testing Centre and direct field sampling. The reflectance of samples was measured using a FieldSpec 3 diffuse reflectance spectrometer. Outliers and sample relation were evaluated using principal component analysis (PCA) and models were developed through partial least square regression (PLSR). For nine watersheds sampled, 20% of the samples were set aside to test prediction and 80% were used to develop calibration models. Depending on the number of samples per watershed, cross validation or independent validation were used. The stability of models was evaluated using coefficient of determination ( $R^2$ ), root mean square error (RMSE), and the ratio performance deviation (RPD). The  $R^2$  (%), RMSE (%), and RPD, respectively, for validation were Anjeni (88, 0.44, 3.05), Bale (86, 0.52, 2.7), Basketo (89, 0.57, 3.0), Benishangul (91, 0.30, 3.4), Kersa (82, 0.44, 2.4), Kola tembien (75, 0.44, 1.9), Maybar (84, 0.57, 2.5), Megech (85, 0.15, 2.6), and Wondo Genet (86, 0.52, 2.7) indicating that the models were stable. Models performed better for areas with high SOC values than areas with lower SOC values. Overall, soil spectroscopy performance ranged from very good to good.

## 1. Introduction

Ethiopia is one of the largest countries in Africa. The vast majority of the society (~80%) depends on agriculture which contributes around 42% of the growth domestic product (GDP) [1]. Despite its strategic importance for the country's economic development, the agricultural sector suffers from low efficiency, population pressure, ineffective land management, and unfavourable land use practices leading to widespread land degradation. Still, the country is strongly committed to environmental protection, rehabilitation, and sustainable land management as evident from Article 92 of its constitution [2], and various internationally supported initiatives have been launched aiming to stop and reverse this trend. Among others, the Growth and Transformation Plan (GTP) is the major one. To address these challenges, efficient and affordable land management practices are required in all

regions of the country and their effectiveness should be analysed and monitored. Alongside, there is pronounced need to improve knowledge on soil resources and to collect reliable data on soil's state and dynamics, allowing for operational assessment and monitoring of this important resource more precisely [3]. Such information is critically important for research and development interventions.

Traditionally, such information was collected through comparatively expensive and slow wet chemistry analysis methods. Every year, many soil samples are collected and analysed for determination of soil properties critical to soil management and crop husbandry using the conventional (wet chemistry) laboratory methods. However, the country is facing a critical challenge to continue with conventional methods as the cost for chemicals is very high and the efficiency of conventional laboratories is low. Soil spectroscopy as an alternative approach has yet to be established in

Ethiopia; therefore, conventional methods of soil analysis currently remain the only available option. The advent of spectroscopy in soil science provided a promising prospect and is nowadays successfully applied in many parts of the world since its early beginnings [3]. Soil spectroscopy is an analytical technique used for soil analysis [4]. It is fast and efficient in producing comprehensive soil information in a short span of time [3–7]. Many soil attributes can be determined with a single scan [4, 6, 8]. Moreover soil spectroscopy requires only a small amount of soil; it is a nondestructive procedure allowing samples to be stored and reanalysed. Finally, it requires less complex infrastructure as compared to the conventional soil analysis. Despite the aforementioned benefits, the technology is not yet widely used in Ethiopia. To our knowledge, the only scientific report on soil spectroscopy in Ethiopia was published by Vågen et al. [9]. In view of the country's urgent need for more efficient technologies, a study on the applicability of soil spectroscopy for the prediction of SOC was conducted in the highlands of Ethiopia. Soil organic carbon is an important indicator for soil health and fertility, serving alongside as a measure for the potential of carbon storage in the soils in relation to climate change mitigation [10]. Therefore, the objective of the study was to evaluate the feasibility of VNIR soil spectroscopy for predicting SOC in the highlands of Ethiopia.

## 2. Materials and Methods

**2.1. Study Area.** The chosen study sites cover most of the agroecological zones, climate conditions, geologies, and soils of the Ethiopian highlands (Figure 1 and Table 1).

The study sites including Bale Mountain, Basketo, Maybar, and Wondogenet are characterised by bimodal type of rainfall, which is manifested in the eastern, northeastern, and south-eastern parts of the country. The other study sites have one prolonged rainy season [11]. The general characteristics of the study sites are given in Table 1. The southern and western parts of the country are mostly under perennial crops; the northern parts of the country are dominated by cereal crops. The farming systems are characterised by small-scale crop-livestock mixed types. In all of the study sites, sustainable land management practices have been introduced in the course of land rehabilitation and reclamation campaigns.

**2.2. Soil Sample Sets.** For this study, a total of 1159 soil samples were chemically and spectrally analysed. The major share of soil samples (713) was acquired from the soil archive of Ethiopia's National Soil Testing Centre (NSTC), representing the soils of the dominant highland areas of the country. We explored the soil archive for the existing soil samples together with the analysed soil parameters, method of analysis, year of sampling, and origin of the samples. Then, sites for this study were selected based on availability of organic carbon data and their representation to the highlands of Ethiopia. Nevertheless, it was not possible to get samples representing the whole range of the highlands collected during the same year and hence we took samples collected from 2006 to 2010 as follows: Benishangul in 2006; Megech in 2008; Bale

and Wondogenet in 2009; and Kola tembien, Kersa, and Basketo in 2010. In addition, 446 soil samples were collected in October–December 2010 from Anjeni and Maybar sites.

Soil samples in the catchments of Anjeni and Maybar (Figure 1) were collected according to a field sampling procedure presented in Amare et al. [12]. The sampling design was optimized to account for the spatial variability of soil properties due to land use and land management activities, as well as for pedogenetic factors varying across spatial and temporal scales. Samples from all other sites analysed and stored at the NSTC were collected following similar sampling designs as outlined previously, addressing soil characterisation and effects of land use and management practices in the watersheds.

**2.3. Chemical Soil Analysis.** The soil organic carbon content of all samples from the NSTC was determined using the wet oxidation Walkley-Black method [13], a standard procedure in Ethiopia. The additionally collected samples from Anjeni and Maybar were dried under shadow conditions, ground, and sieved to pass through 2 mm sieve and then soil organic carbon was determined similarly to the other sample sets. All samples were analysed in the same laboratory following the same procedure to errors due to laboratory procedure and helped to work with a homogeneous sample set.

**2.4. Spectral Measurement.** Pretreatment and spectral measurements were carried out in line with the protocol developed by the global soil spectroscopy group [14, 15]. The reflectances of all soil samples were measured at Ethiopia's National Soil Testing Centre in Addis Abeba, February 2011, using a FieldSpec 3 diffuse reflectance spectrometer (Analytical Spectral Devices, Boulder, CO, USA). The spectra were measured from 350 to 2500 nm at 1 nm interval. A high-intensity mug light source was used to illuminate samples from the bottom through borosilicate Duran glass petri dishes. Roughly, 1 cm layer of soil was poured into the sample holders. A panel consisting of spectralon (Labsphere, North Sutton, USA) was used as a white reference standard between each measurement. The spectrometer was optimised regularly after measuring 10 samples and between shifting sets of different soil samples. Two replicate spectra with 90-degree rotation were collected in order to increase the precision of the measurements. Whenever high variations occurred between the scans of the same sample, the measurement was taken again. After checking the spectral signatures of each repeated measurement, average spectra of the two readings were computed. Spectral inconsistencies (splices) were observed for all samples at 1000 and 1830 nm and were corrected by applying an offset. In order to reduce data and the processing time, only every 10th reading starting from 380 nm was kept for further analysis [14]. Spectral regions below 380 nm or above 2450 nm were removed because they were affected by noise [3].

**2.5. Data Analysis and Model Development.** The geological background, soil type, and agro-ecology of the study areas are diverse (Table 1 and Figure 1). Not only between, but

TABLE 1: Overview description of the study sites.

Sampling site	Agroecological zone*	Rainfall (mm)	Temperature range (°C)**	Dominant soil types***	Geological background
Anjeni	Wet Weyna Dega	1650	9–23 (16)	Alisols, Nitisols, Cambisols, Regosols, Acrisols, Luvisols, and Leptosols	Tertiary Basalts
Bale mountains	Dega-High Dega	870–1064	8.6–15.2 (13.1)	Andosols	Volcanic rocks with volcanic ash material
Basketo	Weyna Dega	~1300	12.8–27.7 (18)	Nitisols and Andosols	Ignimbrites and Rhyolites
Benishangul	Weyna Dega/Kolla	1116	9.5–34.4 (25)	Nitisols	Tertiary Gneisses and Basalts
Kersa	Dega/Weyna Dega	~1400	~18–25 (19)	Vertisols and Nitisols	Mostly tertiary basaltic rocks
Kolatembien	Weyna Dega	600	~16–20 (18)	Regosols, Cambisols, and Leptosols	Antalo Limestones
Maybar	Moist Weyna Dega/moist Dega	1150	12.3–21.4 (16.4)	Phaeozems, Gleysols, Fluvisols, and Leptosols	Alkali-olive Basalts and Tuffs
Megech	Weyna Dega	700–1160	14.2–26.9 (20)	Vertisols, Fluvisols, and Cambisol	Pleistocene to recent alluvial deposits
Wondogenet	Weyna Dega	948	12.6–27.3 (19)	Luvisols, Andosols, and Nitisols	Ignimbrites, Basalts, and Tuffs

\* Hurni, 1998 [11], \*\* the mean annual values are in the parentheses, \*\*\* based on FAO-Unesco soil classification.

also within a given watershed, environmental conditions are highly variable (Table 2). Due to this fact, a separate modelling strategy was followed in this study, developing one model for each watershed. According to the main objectives of this study, which was to test the applicability of soil spectroscopy for various environmental settings in Ethiopian highlands, this choice was further motivated by findings and discussions given by [16, 17]. For exploratory data analysis, the *R* statistics language [18] was used while modelling was done in Unscrambler 10.1 (CAMO, Oslo, Norway).

The dataset was screened for outliers using PCA as well as model residuals. One sample from Bale, one sample from Megech, and three samples from Wondogenet study sites were flagged for removal due to high residual variance as well as high leverage and score distance; we speculate that sample handling errors or fundamentally different soil characteristics (out-of-population sample) could be possible reasons for these variations.

Various spectral preprocessing options were tested, and finally preprocessing with first derivative was found to improve the performance of the models best in terms of stability and interpretability, except for the Bale case study site, where second-order derivative was scoring higher. Derivative is a widely used preprocessing technique [16], which has a baseline correction effect and is able to enhance weak signals [16, 19]. Before modelling, all spectra further underwent autoscaling (mean-centering and variance-scaling), which is a recommended standard procedure for most cases [20].

The raw spectra displayed in Figure 2 reflect the variability of the VNIR spectra in the sample set exemplarily for the Bale case study site. Basically, the shape of all VNIR spectra was similar, showing a steep ascent from 400 to 750 nm, which is characteristic for iron oxides [16]. The

dominant absorption regions (reflectance minima) around 1450 and 1900 nm are usually attributable to  $\text{OH}^-$  and  $\text{H}_2\text{O}$ , masking most of other signals [16]. The higher similarity of the spectra of the derivative spectra given in the lower part of Figure 2 is resulting from the removal of additive and multiplicative scatter effects achieved by the derivative preprocessing techniques.

**2.5.1. Calibration Models.** Partial least square regression (PLSR) was used to develop models for the relationship between soil spectral reflectance and soil organic carbon determined by wet chemistry. The number of samples selected for calibration, validation, and prediction varied from watershed to watershed based on the number of samples. For all watersheds except for Anjeni, models were developed with 80% of the samples with cross validation, while 20% of the samples were randomly selected and set aside for model testing. The distributions of SOC for these samples selected randomly for independent model testing were compared and scrutinized with SOC values samples used for calibration and validation using descriptive statistics (Table 2) and statistical tests discussed later.

**2.5.2. Splitting Samples into Calibration, Validation, and Test Sets.** For Anjeni watershed, the number of samples (302) was sufficient to split into calibration and validation sets. Before splitting the sample sets into calibration and validation sets, 20% of the samples (61 samples) were selected randomly for checking purpose. The 61 samples constitute the test set. After selection and separation of samples for testing set, the remaining 80% of the samples (241 samples) were split again randomly into calibration set (145 samples) and validation set (96 samples). This approach separating

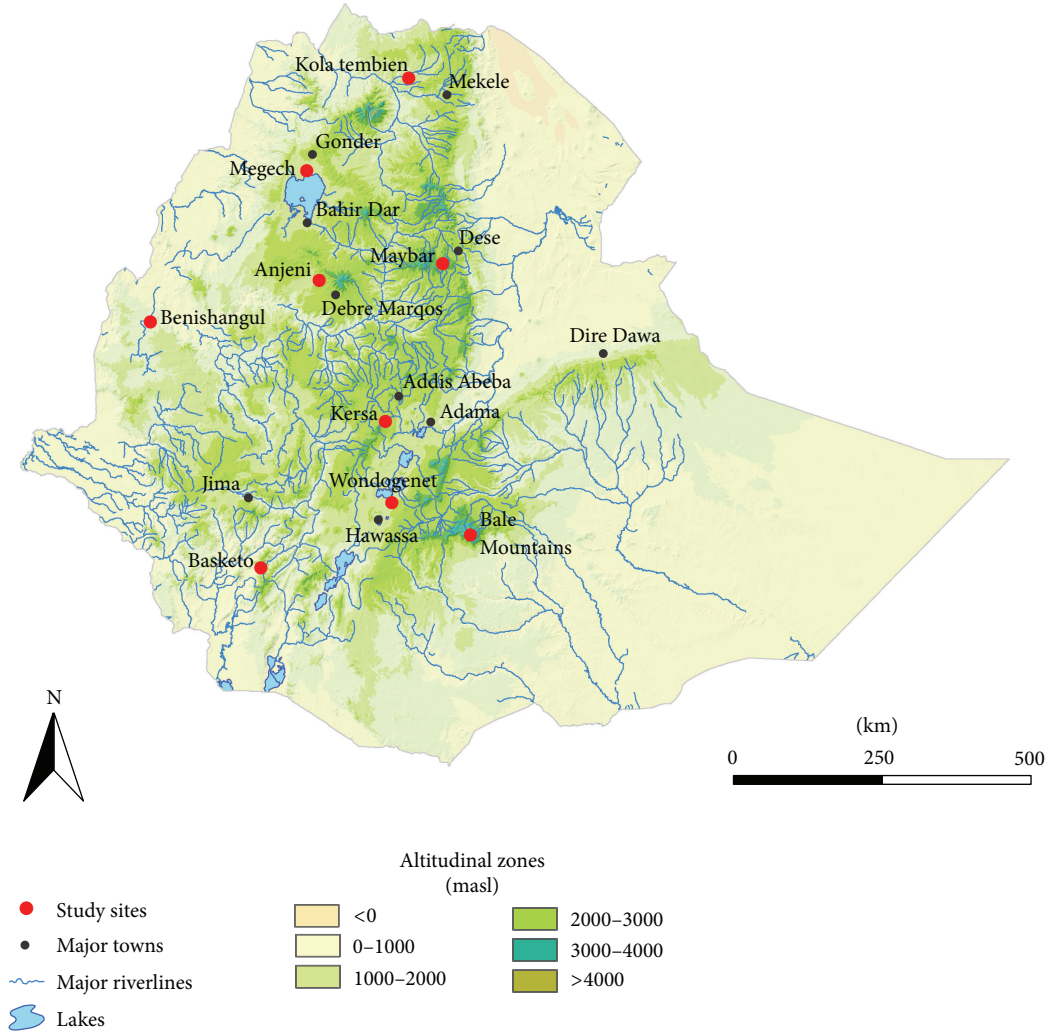


FIGURE 1: Map of Ethiopia indicating the location of the study sites.

calibration, validation, and testing sets is recommended as a robust procedure by Varmuza and Filzmoser [21].

Therefore, the separation of all testing sets was done following the same procedure as given previously. The sample selection of calibration set, validation set, and test set for a given site was based on the size of the samples sets available per study area. The statistical appropriateness of the sample split was scrutinized using the two samples *t*-test for mean homogeneity, Wilcoxon-sum rank test for distribution homogeneity, and the Levene's test for variance homogeneity. Furthermore, descriptive statistics were calculated for total samples, calibration, validation, and test sets as presented in Table 2.

The statistical parameters ( $R^2$ , RMSE, and RPD) were calculated as indicated below.

The coefficient of determination ( $R^2$ ) is as follows:

$$R^2 = \frac{\sum_{i=1}^N (X - \bar{X})(Y - \bar{Y})^2}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}, \quad (1)$$

where  $X$  is the laboratory measured soil organic carbon (%) for each observation and  $\bar{X}$  is the mean value of the laboratory measured while  $Y$  is the predicted soil organic carbon (%) and  $\bar{Y}$  its mean value.

The root mean square error (RMSE) is as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^N \left( \frac{Y - X}{N} \right)^2}, \quad (2)$$

where  $Y$  is the predicted (fitted) value,  $X$  the measured value, and  $N$  is the number of observation.

The ratio of performance deviation (RPD) [22] is as follows:

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}}, \quad (3)$$

where SD is the standard deviation of the measured samples.

Finally, calibration models were evaluated based on coefficient of determination ( $R^2$ ), root mean square error (RMSE), and ratio of performance deviation (RPD) as discussed later in the results part of this paper.



TABLE 2: Statistical summary of the samples analyzed for SOC content (%) for calibration, validation, and test sample sets.

Study sites	Number of samples	Min.	Max.	Med.	Mean	SDV	1st quartile	3rd quartile
<b>Anjeni</b>								
Total	302	0.20	13.68	1.53	1.67	1.34	0.90	2.15
Calibration	145	0.20	13.68	1.39	1.54	1.36	0.84	1.94
Validation	96	0.21	9.90	1.55	1.70	1.27	0.92	2.22
Test	61	0.27	9.74	1.76	1.93	1.40	1.08	2.39
<b>Bale</b>								
Total	74	1.23	7.50	3.14	3.35	1.49	2.26	4.22
Calibration	59	1.23	7.20	3.08	3.25	1.41	2.20	4.22
Test	15	1.84	7.5	3.36	3.78	1.81	2.56	4.47
<b>Basketo</b>								
Total	102	0.28	8.05	2.18	2.30	1.60	0.94	2.77
Calibration	81	0.28	8.05	2.23	2.36	1.69	0.89	2.81
Test	21	0.39	5.6	2.18	2.13	1.23	1.52	2.4
<b>Benishangul</b>								
Total	78	0.10	5.46	0.70	1.19	0.98	0.49	1.80
Calibration	64	0.10	5.46	0.72	1.19	1.01	0.44	1.73
Test	14	0.20	2.87	0.67	1.18	0.90	0.53	1.92
<b>Kersa</b>								
Total	171	0.11	6.07	1.69	1.86	1.02	1.23	2.22
Calibration	135	0.11	6.07	1.77	1.91	1.05	1.24	2.36
Test	36	0.20	4.42	1.63	1.70	0.89	1.20	2.05
<b>Kolatembien</b>								
Total	65	0.28	4.59	0.83	1.10	0.86	0.53	1.29
Calibration	51	0.28	4.59	0.81	1.05	0.85	0.48	1.21
Test	14	0.46	3.90	1.09	1.29	0.93	0.61	1.51
<b>Maybar</b>								
Total	144	0.19	7.58	1.56	1.97	1.33	1.14	2.40
Calibration	115	0.19	7.58	1.54	1.99	1.42	1.09	2.34
Test	29	0.68	4.27	1.71	1.89	0.93	1.35	2.40
<b>Megech</b>								
Total	147	0.16	2.54	0.88	0.92	0.38	0.74	1.11
Calibration	118	0.16	2.54	0.89	0.93	0.39	0.73	1.11
Test	29	0.32	1.44	0.84	0.87	0.29	0.78	1.02
<b>Wondogenet</b>								
Total	71	1.23	8.29	3.1	3.37	1.38	2.34	4.21
Calibration	57	1.23	8.29	3.28	3.36	1.39	2.44	4.02
Test	14	1.56	6.16	2.96	3.36	1.43	2.34	4.30

SDV: the standard deviation, Min: minimum, Max: maximum, Med: median. All splits of the sample sets given in Table 2 have been tested for statistical conformity. For all splits, the test results confirmed the null hypothesis which is assuming homogeneity of the parameters tested.

### 3. Results and Discussion

**3.1. Soil Organic Carbon Distributions in the Study Sites.** As shown in Table 2, soil organic carbon (from wet chemistry analysis) was highly variable between locations and within each location. The highest variability of soil organic carbon contents was observed at Anjeni with values ranging from 0.2% (soil samples from agricultural fields) to 13.68% (samples collected from the top soil of an old forest belonging to Orthodox Church). The latter value is also the maximum SOC value recorded for the entire sample set used for this study. Such high SOC content results from a combination of high rainfall, favourable temperature, and regular turnover of biomass without forest clearing or biomass removal, based on the Ethiopian Orthodox Church's strict rules and regulations for its own forest management. The church's contribution

to and role in forest protection and reforestation have been studied in detail by Wassie Eshete [23]. The highest mean value for soil organic carbon was observed in Wondogenet (3.37%), followed by Bale (3.35%). Both sites are located in high rainfall areas of southern Ethiopia, where agroforestry-based farming systems with generally higher SOC values is dominating. The Bale site is located at the Bale Mountain National Park. The lowest mean values for SOC were found in Megech (0.92%) and Kola tembien (1.10%). Kola tembien is located in the northern and dryer part of the country, where many areas are affected by land degradation and low SOC values. As such, our SOC content analysis findings listed in Table 2 are consistent with those from earlier soil-related studies conducted in the respective parts of Ethiopia. In northern Ethiopia, Mekuria et al. [24], Van de Wauw et al. [25], Abegaz et al. [26], Abegaz and van Keulen [27], and

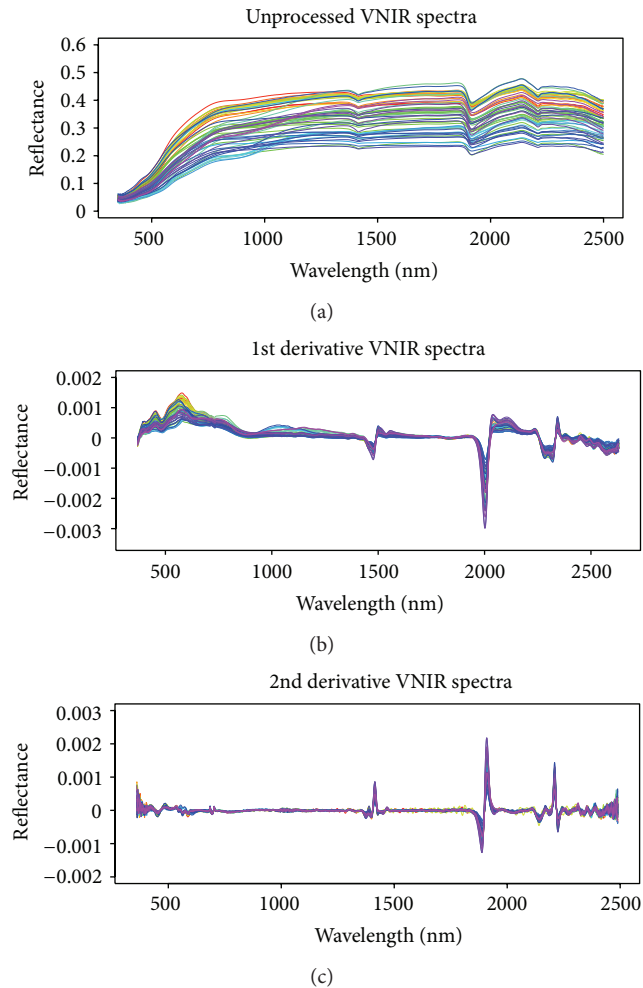


FIGURE 2: Spectral preprocessing from raw spectral data to first derivative and second derivatives for Bale samples.

Tesfahunegn et al. [28] reported SOC contents within the range of the results given in Table 2. For the southern parts of the country, Solomon et al. [29], Lemenih et al. [30], Teklay et al. [31], and Ashagrie et al. [32] found similar carbon contents. Yimer et al. [33] and Yimer et al. [34] in the Bale Mountains, while Spaccini et al. [35] and Tulema et al. [36] reported similar results for the central and northeastern highlands. Based on these findings, we conclude that the results presented in Table 2 are adequately well representing the respective areas in terms of the SOC range covered.

**3.2. Spectral Data Analysis Using PCA.** As described previously, the spectral readings from all study sites were analysed using PCA in order to detect outliers and patterns in the data; however, this was also done to detect clusters or subgroups for aggregated model building, comparing the sites to each other. Given the high variability listed in Table 1, PCA seems to be a promising and powerful tool for reducing the amount of redundant data and dimension reduction. Therefore, PC1 and PC2 were calculated from the raw spectra of selected study sites and the scores were displayed in a scatterplot (Figure 3). The cumulative variance explained by the first two

PCs was higher than 90% in all cases. Figure 3 displays the score plots for the study sites of Kersa, Kola tembien, and Maybar, indicating that these watersheds are very different; only little similarity exists between the scores, even if the 95% data coverage ellipse might suggest a small overlap. Similarly, Bale, Kola tembien, and Megech study sites were projected into the PC space and show well-defined regions with few overlapping samples. Samples from Maybar were also separated from samples of Wondogenet and Bale Mountain with only few overlapping. Samples from Megech were also separated from samples of Wondogenet and Bale. While the Megech samples cover rather compact region in the plot, the samples from Kola tembien spread less uniformly over much larger area, which might indicate various levels of complexity and variability within the watershed. In Figure 3 samples Maybar covers compact region with five distinct samples far away from the centre. Still Maybar samples were also separated from Wondogenet and Bale samples with only few overlap. Figure 3 also displays samples from Megech together with Bale and Kersa. This figure clearly shows a certain degree of similarity between Bale and Kersa watersheds, indicating the potential for developing a combined model. The discussed results are not very surprising when one considers the highly diverse environmental conditions present in the different catchments, and the uneven spread of samples the in PCA space of Kola tembien and Bale even suggest to further subset models. Similar results and conclusions have been reported by McDowell et al. [17] and Stenberg et al. [16], stating that subset models score higher in terms of precision and accuracy. The downside of course of subset model is their lack of stability and robustness.

Thus, using PCA, it was possible to reveal the data structure in the samples relating to their sources of origin, with varying levels of overlap. The method is fast due to complete absence of soil analysis by wet chemistry procedure. Odlare et al. [37] have suggested that field characteristics must be significantly different to be distinguished from each other. Similarly, PCA has been used effectively to separate organic and nonorganic wine products [38], soil colors [18], soluble and less soluble elements [39], and sediment sources [40]. The use of PCA also reported to separate soil samples according to management practices such as tillage [41] and composting [42].

**3.3. Model Calibration, Validation, and Testing.** Using PLSR, the results show in general that visible/near-infrared spectroscopy performed well for all different agroecological conditions, soil types, and land management practices of Ethiopia. As shown in Table 3, all values of the coefficient of determination ( $R^2$ ) were in a range of very good to good with references to Stenberg and Rossel [43] and Chang et al. [44].

For all sites except Kola tembien and Kersa, the ratio of performance deviation (RPD) was above 2.5 for validation—a value that is considered very adequate for developing stable prediction models for soil organic carbon [45]. The value of RPD for the present study ranged between 2.79 (Maybar) and 6.04 (Benishangul) for calibration and between 1.92 (Kola tembien) and 3.39 (Benishangul) for the validation set. The

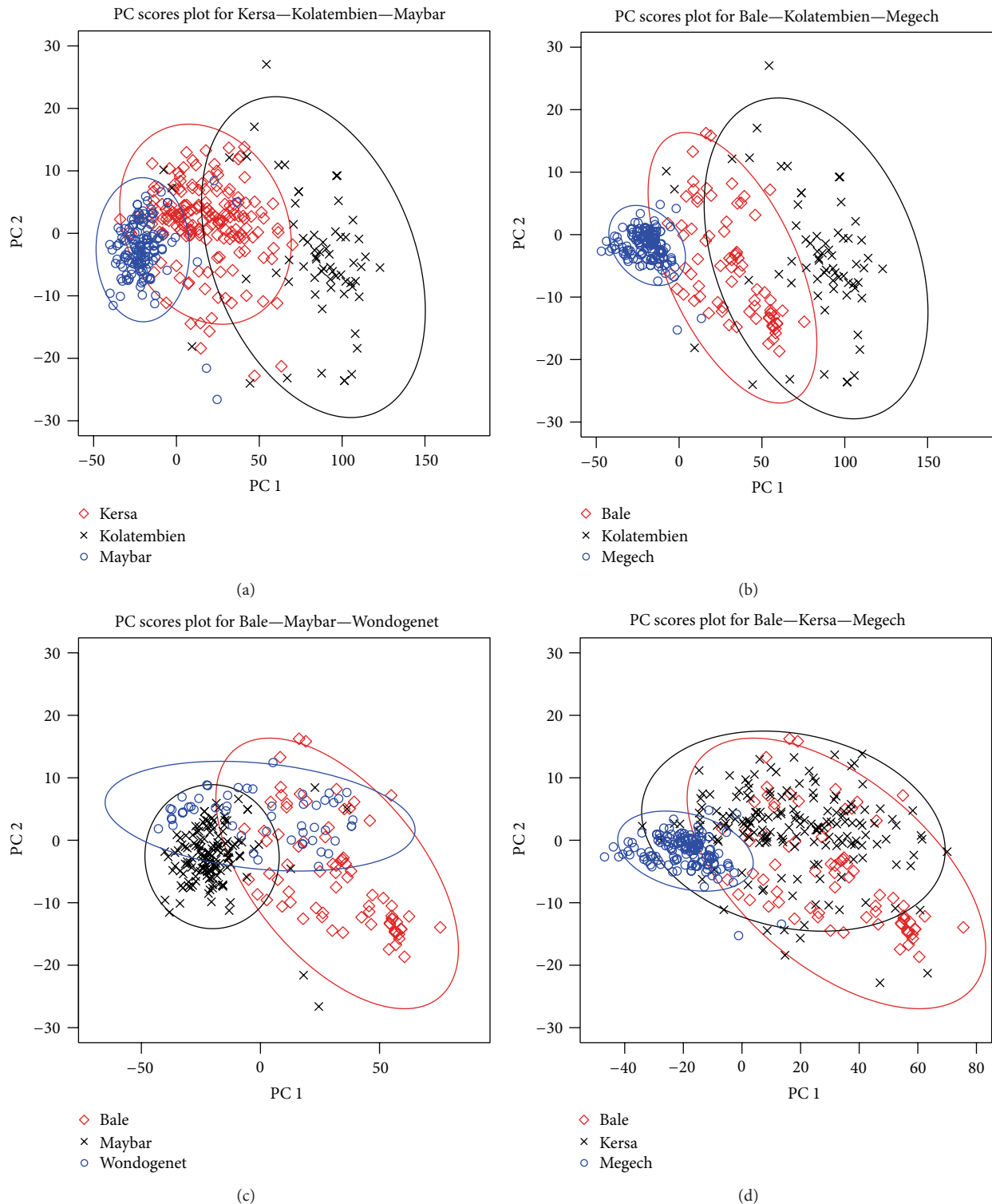


FIGURE 3: PCA scores for the study sites.

comparatively poor performance of the calibration model for Kola tembien may be related to the low levels of SOC, which is related to the high level of soil degradation [46] and high carbonate. This reflects the findings by other authors who reported unstable calibration models for areas coined by low

SOC and high carbonate levels [16]. Benishangul is covered by a single soil type (Nitisols), which developed on Gneisses and Basalts. The area is further dominated by uniform land use (cropland). This low variation of environmental conditions is well expressed by the good model fit. On the other

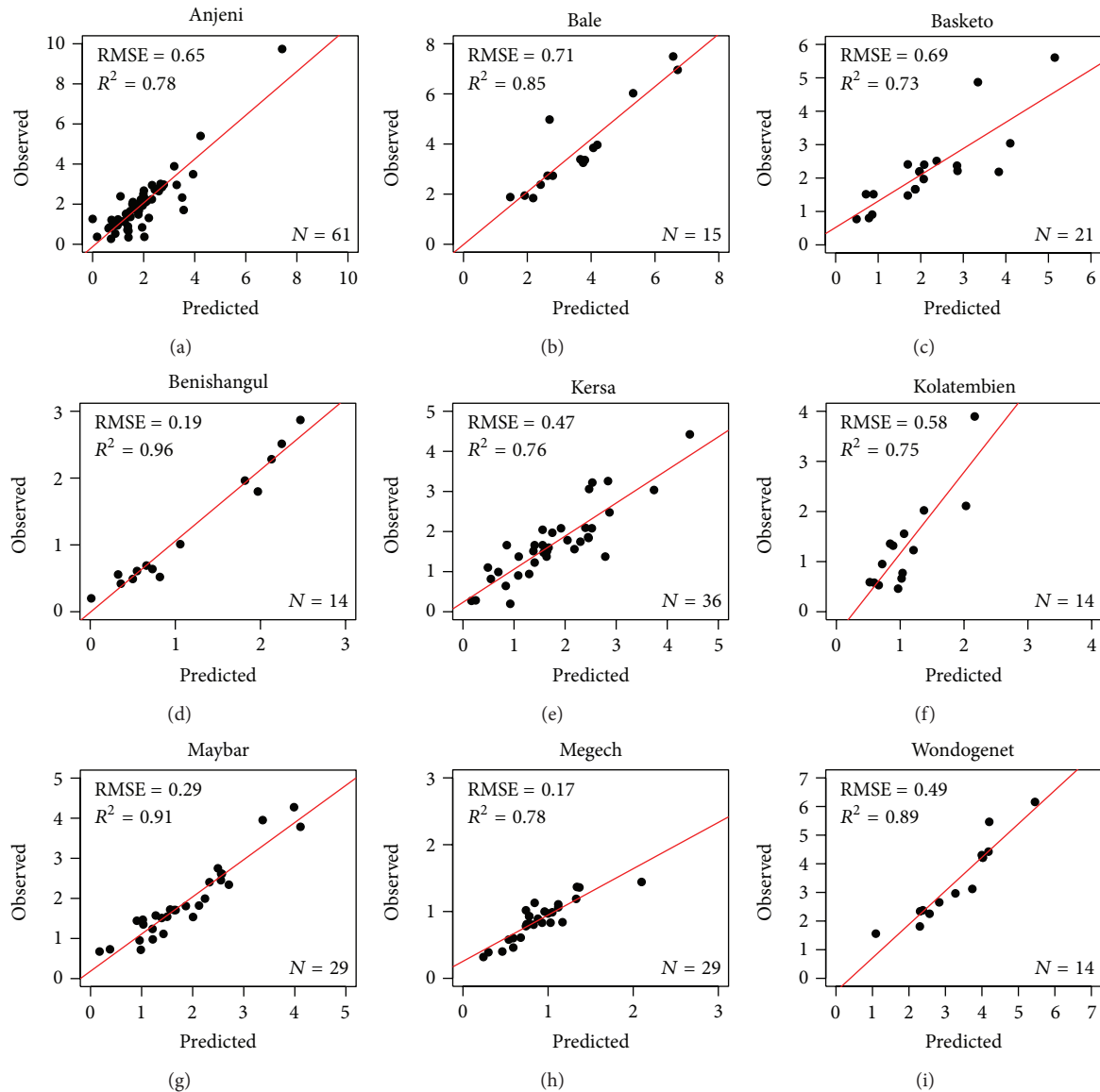


FIGURE 4: Comparison between SOC (%) determined by Walkley-Black method (observed) and spectral model (predicted) for the test sample sets.

hand, Anjeni, watershed in Gojam, provided very acceptable model results despite the high range of SOC (Table 2), high variability of soil types (Table 1), and land uses [47].

The range of the root mean square error (%) varied between 0.11 (Megech) and 0.53 (Maybar) for the calibration and between 0.15 (Megech) and 0.57 (Maybar) for validation. Rossel et al. [45] reported a range of root mean square error (RMSE) comparable to the present results. Rossel and Behrens [48] achieved maximum precision with 0.75% RMSE and 89%  $R^2$  for soil organic carbon, using artificial neural network (ANN) with wavelet coefficients—this is comparable to the present results (Table 3). They reported higher RMSE and lower  $R^2$  with multiple linear regressions (MLR), PLSR, multivariate adaptive regression splines (MARS), support vector machines (SVM), random forests (RF), and boosted trees (BT)—when compared to ANN—indicating that the

results are fairly reliable. After calibration and validation, the developed models were tested with ~20% of the samples from each study area, which is presented in Figure 4.

#### 4. Conclusions and Recommendations

From the results of the study, it is possible to conclude that soil spectroscopy is an effective method for predicting soil organic carbon in the highlands of Ethiopia. There are considerable variations in the performance of the local models, which are mostly linked to the variability of the environmental settings prevailing within the watersheds. Models for the study sites in the moist southern and western parts of the country seem to perform slightly better. In general terms, the relationship between predicted and measured values of soil organic carbon was high for all sites, regardless of the



TABLE 3: Summary of model results.

	Model	Number of samples	$R^2$	RMSE (%)	RPD
Anjeni	Calibration	145	90	0.46	3.1
	Validation	96	88	0.44	3.1
Bale	Calibration	59	93	0.42	3.7
	Validation	59	86	0.52	2.7
Basketo	Calibration	81	93	0.46	3.9
	Validation	81	89	0.57	3.0
Benishangul	Calibration	64	97	0.18	6.0
	Validation	64	91	0.30	3.4
Kersa	Calibration	135	90	0.35	3.2
	Validation	135	82	0.44	2.4
Kolatembien	Calibration	51	88	0.33	2.3
	Validation	51	75	0.44	1.9
Maybar	Calibration	115	87	0.53	2.8
	Validation	115	84	0.57	2.5
Megech	Calibration	118	94	0.11	4.0
	Validation	118	85	0.15	2.6
Wondogenet	Calibration	57	90	0.46	3.2
	Validation	57	86	0.52	2.7

$R^2$ : the coefficient of determination, RMSE: root mean square error, RPD: ratio of performance deviation.

high variability of soil organic carbon within watersheds like Anjeni. The RPD values for validation ranged from 1.9 (Kola tembien) to 3.4 (Benishangul). This indicates fairly stable models and thus the models developed from these data could be used to build local spectral libraries with sufficient predictive power for local land use and management-related applications. Nevertheless, there is a potential for further improvement of the presented local models, which could be achieved by enlarging the range of the sample set and adding more samples from the respective sites, which would also allow for developing robust calibration models.

Considering the results from the PCA, there is a potential for research in investigating the performance of aggregated models and other modelling strategies, alongside with the use of nonparametric calibration methods as suggested by many authors. While the present study only focussed on soil organic carbon, soil spectroscopy's ability to predict other chemical, biological, and physical soil parameters in Ethiopia remains to be explored.

## Acknowledgments

This paper is based on the work conducted within the framework of Swiss National Centre of Competence in Research (NCCR) North South: Research Partnerships for Mitigating Syndromes of Global Change. The authors wish to thank the Centre for Development and Environment (CDE), University of Bern, for providing the spectrometer. They would also like to express thier gratitude to the Ethiopian National Soil Testing Centre for providing access to the soil archive. Thanks are due to the Amhara Regional Agricultural

Research Institute (ARARI) for its support and facilitation. They further extend their gratitude to Berhanu Debele and Deresse Gebrewold for their facilitation and administrative support. Finally, thanks are due to Lorenz Ruth for his keen assistance with spectral measurement.

## References

- [1] L. Barrow, P. Mwanakatwe, H. Hashi, H. Zerihun, and V. Ndisale, "Federal Democratic Republic of Ethiopia Country Strategy Paper," 2011.
- [2] Ethiopia, *The Constitution of the Federal Democratic Republic of Ethiopia*, Addis Abeba, Ethiopia, 1995.
- [3] K. D. Shepherd and M. G. Walsh, "Development of reflectance spectral libraries for characterization of soil properties," *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 988–998, 2002.
- [4] K. D. Shepherd and M. G. Walsh, "Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries," *Journal of Near Infrared Spectroscopy*, vol. 15, no. 1, pp. 1–19, 2007.
- [5] D. J. Brown, K. D. Shepherd, M. G. Walsh, M. Dewaynemays, and T. G. Reinsch, "Global soil characterization with VNIR diffuse reflectance spectroscopy," *Geoderma*, vol. 132, no. 3–4, pp. 273–290, 2006.
- [6] R. A. V. Rossel, D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad, "Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties," *Geoderma*, vol. 131, no. 1–2, pp. 59–75, 2006.
- [7] D. Brunet, B. G. Barthès, J.-L. Chotte, and C. Feller, "Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: effects of sample grinding and set heterogeneity," *Geoderma*, vol. 139, no. 1–2, pp. 106–117, 2007.
- [8] E. Ben-Dor and A. Banin, "Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties," *Soil Science Society of America Journal*, vol. 59, no. 2, pp. 364–372, 1995.
- [9] T. V. Vågen, L. A. Winowiecki, A. Abegaz, and K. M. Hadgu, "Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia," *Remote Sensing of Environment*, vol. 134, pp. 266–275, 2013.
- [10] R. T. Conant, S. M. Ogle, E. A. Paul, and K. Paustian, "Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation," *Frontiers in Ecology and the Environment*, vol. 9, no. 3, pp. 169–173, 2011.
- [11] H. Hurni, *Agroecological Belts of Ethiopia*, University of Berne, Bern, Switzerland, 1998.
- [12] T. Amare, A. Terefe, Y. G. Selassie, B. Yitaferu, B. Wolfgramm, and H. Hurni, "Soil properties and crop yields along the terraces and toposequece of Anjeni Watershed, Central Highlands of Ethiopia," *Journal of Agricultural Science*, vol. 5, no. 2, 2013.
- [13] A. Walkley and A. Black, "An examination of degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method," *Soil Science*, vol. 37, pp. 29–37, 1934.
- [14] R. V. Rossel, "The soil spectroscopy group and the development of a global soil spectral library," in *Proceedings of the European*

- Geosciences Union (EGU) General Assembly Conference*, vol. 11, abstracts 14021, April 2009.
- [15] J. Wetterlind, B. Stenberg, and R. A. V. Rossel, "Soil analysis using visible and near infrared spectroscopy," in *Plant Mineral Nutrients: Methods and Protocols, Methods in Molecular Biology*, F. J. M. Maathuis, Ed., 1st edition, 2013.
  - [16] B. Stenberg, R. A. V. Rossel, A. M. Mouazen, and J. Wetterlind, "Visible and near infrared spectroscopy in soil science," in *Advance in Agronomy*, vol. 107, Chapter 5, pp. 163–215, Academic Press, New York, NY, USA, 2010.
  - [17] M. L. McDowell, G. L. Bruland, J. L. Deenik, and S. Grunwald, "Effects of subsetting by carbon content, soil order, and spectral classification on prediction of soil total carbon with diffuse reflectance spectroscopy," *Applied and Environmental Soil Science*, vol. 2012, Article ID 294121, 14 pages, 2012.
  - [18] A. M. Mouazen, R. Karoui, J. Deckers, J. De Baerdemaeker, and H. Ramon, "Potential of visible and near-infrared spectroscopy to derive colour groups utilising the Munsell soil colour charts," *Biosystems Engineering*, vol. 97, no. 2, pp. 131–143, 2007.
  - [19] R. A. V. Rossel, "ParLeS: software for chemometric analysis of spectroscopic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 72–83, 2008.
  - [20] M. K. Boysworth and K. S. Booksh, "Aspects of Multivariate Calibration Applied to Near-Infrared Spectroscopy," in *Handbook of Near-Infrared Analysis*, D. A. Burns and E. W. Ciurczak, Eds., pp. 207–2228, CRC Press, New York, NY, USA, 3rd edition, 2008.
  - [21] K. Varmuza and P. Filzmoser, *Introduction To Multivariate Statistical Analysis in Chemometrics*, vol. 64, no. 4, CRC Press, New York, NY, USA, 2009.
  - [22] P. C. Williams, *Near Infrared Technology in the Agriculture and Food Industries*, P. Williams and K. Norris, Eds., American Association of Cereal Chemists, St. Paul, Minn, USA, 1st edition, 1987.
  - [23] A. Wassie Eshete, *Ethiopian Church Forests? Opportunities and Challenges for Restoration*, Wageningen University, The Netherlands, 2007.
  - [24] W. Mekuria, E. Veldkamp, M. Haile, J. Nyssen, B. Muys, and K. Gebrehiwot, "Effectiveness of exclosures to restore degraded soils as a result of overgrazing in Tigray, Ethiopia," *Journal of Arid Environments*, vol. 69, no. 2, pp. 270–284, 2007.
  - [25] J. Van de Wauw, G. Baert, J. Moeyersons et al., "Soil-landscape relationships in the basalt-dominated highlands of Tigray, Ethiopia," *Catena*, vol. 75, no. 1, pp. 117–127, 2008.
  - [26] A. Abegaz, H. van Keulen, and S. J. Oosting, "Feed resources, livestock production and soil carbon dynamics in Teghane, Northern Highlands of Ethiopia," *Agricultural Systems*, vol. 94, no. 2, pp. 391–404, 2007.
  - [27] A. Abegaz and H. van Keulen, "Modelling soil nutrient dynamics under alternative farm management practices in the Northern Highlands of Ethiopia," *Soil and Tillage Research*, vol. 103, no. 2, pp. 203–215, 2009.
  - [28] G. B. Tesfahunegn, L. Tamene, and P. L. G. Vlek, "Catchment-scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia," *Soil and Tillage Research*, vol. 117, pp. 124–139, 2011.
  - [29] D. Solomon, J. Lehmann, T. Mamo, F. Fritzsche, and W. Zech, "Phosphorus forms and dynamics as influenced by land use changes in the sub-humid Ethiopian highlands," *Geoderma*, vol. 105, no. 1-2, pp. 21–48, 2002.
  - [30] M. Lemenih, E. Karlun, and M. Olsson, "Soil organic matter dynamics after deforestation along a farm field chronosequence in southern highlands of Ethiopia," *Agriculture, Ecosystems and Environment*, vol. 109, no. 1-2, pp. 9–19, 2005.
  - [31] T. Teklay, A. Nordgren, and A. Malmer, "Soil respiration characteristics of tropical soils from agricultural and forestry land-uses at Wondo Genet (Ethiopia) in response to C, N and P amendments," *Soil Biology and Biochemistry*, vol. 38, no. 1, pp. 125–133, 2006.
  - [32] Y. Ashagrie, W. Zech, G. Guggenberger, and T. Mamo, "Soil aggregation, and total and particulate organic matter following conversion of native forests to continuous cultivation in Ethiopia," *Soil and Tillage Research*, vol. 94, no. 1, pp. 101–108, 2007.
  - [33] F. Yimer, S. Ledin, and A. Abdelkadir, "Concentrations of exchangeable bases and cation exchange capacity in soils of cropland, grazing and forest in the Bale Mountains, Ethiopia," *Forest Ecology and Management*, vol. 256, no. 6, pp. 1298–1302, 2008.
  - [34] F. Yimer, S. Ledin, and A. Abdelkadir, "Changes in soil organic carbon and total nitrogen contents in three adjacent land use types in the Bale Mountains, south-eastern highlands of Ethiopia," *Forest Ecology and Management*, vol. 242, no. 2-3, pp. 337–342, 2007.
  - [35] R. Spaccini, J. S. C. Mbagwu, P. Conte, and A. Piccolo, "Changes of humic substances characteristics from forested to cultivated soils in Ethiopia," *Geoderma*, vol. 132, no. 1-2, pp. 9–19, 2006.
  - [36] B. Tulema, J. B. Aune, F. H. Johnsen, and B. Vanlauwe, "The prospects of reduced tillage in tef (*Eragrostis tef* Zucca) in Gare Arera, West Shawa Zone of Oromiya, Ethiopia," *Soil and Tillage Research*, vol. 99, no. 1, pp. 58–65, 2008.
  - [37] M. Odlare, K. Svensson, and M. Pell, "Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field," *Geoderma*, vol. 126, no. 3-4, pp. 193–202, 2005.
  - [38] D. Cozzolino, M. Holdstock, R. G. Damberg, W. U. Cynkar, and P. A. Smith, "Mid infrared spectroscopy and multivariate analysis: a tool to discriminate between organic and non-organic wines grown in Australia," *Food Chemistry*, vol. 116, no. 3, pp. 761–765, 2009.
  - [39] S. Xiong, Y. Zhang, Y. Zhuo, T. Lestander, and P. Geladi, "Variations in fuel characteristics of corn (*Zea mays*) stovers: general spatial patterns and relationships to soil properties," *Renewable Energy*, vol. 35, no. 6, pp. 1185–1191, 2010.
  - [40] G. Alaoui, M. N. Léger, J.-P. Gagné, and L. Tremblay, "Assessment of estuarine sediment and sedimentary organic matter properties by infrared reflectance spectroscopy," *Chemical Geology*, vol. 286, no. 3-4, pp. 290–300, 2011.
  - [41] B. Govaerts, K. D. Sayre, and J. Deckers, "A minimum data set for soil quality assessment of wheat and maize cropping in the highlands of Mexico," *Soil and Tillage Research*, vol. 87, no. 2, pp. 163–174, 2006.
  - [42] M. V. Gil, L. F. Calvo, D. Blanco, and M. E. Sánchez, "Assessing the agronomic and environmental effects of the application of cattle manure compost on soil by multivariate methods," *Bioresource Technology*, vol. 99, no. 13, pp. 5763–5772, 2008.
  - [43] B. Stenberg and R. A. V. Rossel, "Diffuse reflectance spectroscopy for high-resolution soil sensing," in *Proximal Soil Sensing*, R. A. V. Rossel, A. B. McBratney, and B. Minasny, Eds., vol. 1 of *Progress in Soil Science*, pp. 29–47, Springer, The Netherlands, 2010.

- [44] C.-W. Chang, D. A. Laird, M. J. Mausbach, and J. Hurburgh C.R., "Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties," *Soil Science Society of America Journal*, vol. 65, no. 2, pp. 480–490, 2001.
- [45] R. A. V. Rossel, R. N. McGlynn, and A. B. McBratney, "Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy," *Geoderma*, vol. 137, no. 1-2, pp. 70–82, 2006.
- [46] K. Vancampenhout, J. Nyssen, D. Gebremichael et al., "Stone bunds for soil conservation in the northern Ethiopian highlands: impacts on soil fertility and crop yield," *Soil and Tillage Research*, vol. 90, no. 1-2, pp. 1–15, 2006.
- [47] W. Alemu, T. Amare, B. Yitaferu, Y. G. Selassie, B. Wolfgramm, and H. Hurni, "Impacts of soil and water conservation on land suitability to crops: the case of Anjeni Watershed, Northwest Ethiopia," *Journal of Agricultural Science*, vol. 5, no. 2, 2013.
- [48] R. A. V. Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1-2, pp. 46–54, 2010.



